AN ADAPTIVE R-ESTIMATE

BY

MICHAEL LEE COHEN

TECHNICAL REPORT NO. 5

DECEMBER 15, 1980

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

AN ADAPTIVE R-ESTIMATE

BY

MICHAEL LEE COHEN

TECHNICAL REPORT NO. 5

DECEMBER 15, 1980

DEPARTMENT OF STATISTICS

STANFORD   UNIVERSITY

STANFORD, CALIFORNIA

AN ADAPTIVE R-ESTIMATE

Michael Lee Cohen, Ph.D.
Stanford University, 1980

In 1962 Hajek proposed a test for location which was uniformly asymptotically fully efficient over a large class of distributions. Van Eeden subsequently derived the asymptotic theory for the corresponding R-estimate. Many authors have expressed reservations with regard to the small sample performance to be expected from this approach. In contrast to the methods of Hajek and Van Eeden, the procedure proposed here uses the entire data set to estimate the score function of the locally most powerful rank test. Using nearest neighbor density estimation methods, an estimate of the score function for the asymptotically most powerful grouped rank test is constructed. This function is itself a step function approximation to the score function for the locally most powerful rank test. Large sample distribution and optimality results are obtained for both the adaptive rank test and the corresponding R-estimate of location. Small sample monte-carlo results are provided.

TABLE OF CONTENTS

1. INTRODUCTION

In 1955, Stein [27] suggested that full asymptotic efficiency in testing and estimation for a large class of distributions could be obtained by utilizing the information about the density contained in the data. This idea was further developed by Hajek [10] and Van Eeden [31]. Hajek demonstrated that by estimating the score function of the locally most powerful rank test, $f'(F^{-1}(u))/f(F^{-1}(u))$ an asymptotically most powerful test of location could be constructed. Using the methods of Hodges and Lehmann [12], Van Eeden utilized Hajek's test to form an asymptotically fully efficient estimate of location. These results are interesting theoretically; but many authors, e.g., Hajek [10], Switzer [29], Huber [15], Wesley [32], and Hogg [13], have expressed reservations concerning the small sample properties achievable through this approach. There are serious questions about the applicability of the results of Hajek and Van Eeden to real world data. In their approach the sample is split into two parts which are used for separate purposes. The first part is a vanishingly small fraction of the data which is used to estimate the score function. The second part of the data determines the ranks to be used with the coefficients based on the first part. This approach raises doubts about the stability of the estimate of the score function for small samples. Beran [5] has partially avoided this difficulty, but his proposed estimate of the score function also appears likely to perform poorly for small samples. No monte carlo

1

results are available concerning the small sample performance of any of the above-mentioned adaptive estimates of location.

The methods we employ differ from those of Hajek and Van Eeden mainly in two respects. First, the entire data set is used to estimate the score function and to perform the resulting test of location. Secondly, by the use of density estimation techniques, an estimate of the score function of the asymptotically most powerful grouped rank test is constructed. This function is itself a step function approximation to the score function of the locally most powerful rank test; and in approximating the score function in this way, it is hoped to gain small sample stability of the estimate.

In the manner of Hodges and Lehmann [12], and following Van Eeden, it is possible to form an estimate of location from the resulting rank test, i.e., an R-estimate. It is hoped that if this two-sample rank test adapts well to the data, the same should be true of the resulting estimate of location. Large sample optimality results are shown for both the proposed adaptive rank test and the associated estimate of location. Finally, small sample monte carlo results are provided.

2. DEVELOPMENT AND HISTORY OF ADAPTIVE RANK TESTS

2.1. The Testing Framework

Assume that we have two samples $(x_1, x_2, \ldots, x_n)$, and $(y_1, y_2, \ldots, y_n)$. Let x_i be i.i.d. $F(x)$, y_j i.i.d. $F(x-d)$, F symmetric, with density f. Let $N = 2n$. Also, let $\underset{\sim}{\omega} = (\omega_1, \omega_2, \ldots, \omega_n)$ be defined by

$$\omega_i = \begin{cases} 0 \text{ if the ith order statistics of the combined sample is an } x \\ 1 \text{ otherwise} \end{cases} .$$

Then a linear rank test is a statistic of the form $\Sigma_{i=1}^{N} c(i/N+1)\omega_i$, where $\{c(i/N+1)\}$ is a vector of constants determined by some function $c(\ )$. A grouped rank test is a statistic with the above form and the added constraint that the number of different $c(i/N+1)$ be finite, i.e., the range of $c(\ )$ is a finite set. Let us assume, as above, that the second sample $\underset{\sim}{Y}$ is d units to the right of the first sample, $\underset{\sim}{X}$. Then to see the ω-vector more clearly one could write down the order statistics of the combined sample and replace every value with 0, or 1, depending on which sample that data point came from. A typical ω-vector with the above shift might look like 00000101100101110111. It is easy to imagine the nature of a desirable vector of constants from the chain of 0's and 1's. One would like a statistic which is large for positive changes in the location of the second sample. The constants for the intermediate ω_i's are relatively unimportant, since the 0's and 1's may be equally represented there. At the extremes, however, the pattern of 0's and 1's

3

depends heavily on the size of the tails of the distribution. For a given displacement a typical ω-vector for a short-tailed distribution might look like 00000001001110111111, whereas a typical ω-vector for a long-tailed distribution might look like 10100001001110111001. The coefficients should take into account this dependence on the size of the tails by attempting to place more weight on the extremes of the ω-vectors for distributions with short tails than for those with long tails.

It is well known, Hajek [10], that for testing the hypothesis that $d > 0$ versus $d = 0$, the locally most powerful rank test (LMPRT) has $c(i/N+1) = J(i/N+1)) = -f'(F^{-1}(I/N+1))/f(F^{-1}(i/N+1))$. We shall consider the possibility of estimating $J(u)$ from the data and using the resulting estimate in a rank test, i.e., adapting. Such a test would hopefully have large-sample optimality properties as well as good small-sample performance for a large family of distributions, and we will verify that such is the case for the proposed test. When one is unsure of the actual distribution of the data, one could consider using such a test, instead of, say, the normal scores test or the sign test, (two-sample tests which are asymptotically most powerful for testing $d > 0$ against $d = 0$ for data which has, respectively, a normal or a Laplace distribution).

2.2. Density Estimation

To estimate functions which are approximations to the score function, it is obvious that estimates of the density, at least, and possibly its derivative need to be considered. There are two main

4

types of density estimates currently in the literature, kernel estimates and nearest neighbor density estimates. Kernel estimates were initially developed by Rosenblass [25] and Parzen [22]. Nearest neighbor density estimates were developed by Fix and Hodges [8], and more formally by Loftsgaarden and Quesenberry [19].

In density estimation we are actually estimating an integral of the density over an interval containing the point at which we wish to determine the density, and then dividing by the length of the interval. This estimates an average of the density in a neighborhood of the point. The differences in the two procedures lie in the determination of the interval. In kernel estimation, we (essentially) preset a band-width and count the number of points in that band around the point. In nearest neighbor estimation, we decide on how many data points we would like to use for each point, here designated as k(N) where N is the size of the sample, and determine the smallest symmetric band around the point which encloses the required number of data points.

We will concentrate on nearest neighbor estimates in the hope of obtaining stable estimation in the tails where the density of observed points is low. Better estimation in the tails will correspond to better estimation of the score function J(u) for u near zero and one. This in turn should lead to better estimation of the coefficients of the rank test for ω's in the crucial regions as described in the previous section. We feel that nearest neighbor estimates should perform better than kernel estimates for small samples because

the size of the window is adapted to the number of data points in the area, instead of being fixed as in kernel estimation.

The results that follow only use properties of density function estimates that are commonly held by both kernel estimates and nearest neighbor estimates. Both types of estimate are location invariant, clearly; since if we add a constant c to the data and estimate the density at $x+c$, we will get the same value as the density estimate at x of the untransformed data. Both estimates are continuous. Also, both estimates converge uniformly with probability one. We add that under general conditions which will not be described here, various convergence results are shown to hold simultaneously for both types of estimate in the work of Moore and Yackel [20]. Therefore, any future investigators could certainly study the procedures to be described here with kernel estimates replacing nearest neighbor estimates everywhere.

2.3. An Adaptive Rank Test Coefficient

It is our objective to construct an estimate of $J(u)$ which is well behaved for small samples. We recall that
$J(u) = -f'(F^{-1}(u))/f(F^{-1}(u))$. There is certainly no difficulty in obtaining an estimate $\hat{F}^{-1}$ of F^{-1}. The most commonly used estimate is
$\hat{F}^{-1}(u) = X_{([nu]+1)}$, where $X_{(i)}$ represents the i-th order statistic from a sample of n i.i.d. X's. For this estimate, under the condition of absolute continuity, we have convergence with probability one by proposition (i), page 423, Rao [24]. There exist in the literature several estimates of both f' and f, but these tend not to be

6

well-behaved for small samples, especially the estimates of f'. It probably is not a good idea to divide by these unstable estimates so that direct substitution in the formula for J(u) will very likely do poorly. However, noticing that

$$J(u) = \frac{d}{du} - f(F^{-1}(u)) = -\lim_{h \to 0} \frac{f(F^{-1}(u+h)) - f(F^{-1}(u-h))}{2h} \quad ,$$

we might consider using $\hat{J}(u) = -[\hat{f}(\hat{G}(u+b)) - \hat{f}(\hat{G}(u-b))]/2b$, for some fixed b, where $\hat{f}$ is some estimate of the density, and $G(u) = F^{-1}(u)$. Pursuing this notion, let us proceed as follows:

Let $0 < \lambda_1 < \lambda_2 < \ldots < \lambda_{k+1} = 1$ be any k fractiles (i.e., partitions of [0,1]), and let $P(i) = [N\lambda_i] + 1$. Then define an estimate of J(u) by

$$\hat{L}_n(u) = - [\hat{f}(\hat{G}(\lambda_i)) - \hat{f}(\hat{G}(\lambda_{i-1}))]/(\lambda_i - \lambda_{i-1}) \quad ,$$

for $\lambda_{i-1} < u < \lambda_i$.

This estimate is now identical in functional form to the score function of a test which was proposed by Gastwirth [8]. In this paper, Gastwirth showed that the test with score function

$$L(u) = [f(G(\lambda_{i-1})) - f(G(\lambda_i))]/(\lambda_i - \lambda_{i-1})$$

generates the asymptotically most powerful grouped rank test based on the fractiles $[\lambda_j]$ when the density underlying the data is f, under certain regularity conditions on f. We remark that a grouped rank test statistic is a statistic of the form $\Sigma_{j=1}^{k+1} c_j (\Sigma_{i=P(j-1)}^{P(j)} \omega_i)$. This collects the ω-vectors' contribution to the test statistic into

7

k+1 groups and gives each observation in any group equal weight.
Thus we may consider the rank test using the proposed estimate
of $J(u)$ as an adaptative asymptotically most powerful grouped rank
test.

Hogg [13] suggested an approach similar to this in his review
of adaptive robust procedures. Quoting, "but I can imagine $J(u)$
being estimated by a curve constructed from a few line segments."
The estimate proposed here may be regarded as a development of this
suggestion. Also, after the present work was begun, Parzen [23]
noted the equality of the score function to a single derivative of
$f(G(u))$ with respect to u and mentioned the possibility of using this
fact in analyzing data.

2.4. Comparison With Previous Work

Hajek [1962], considered the following procedure. The X and Y
samples are randomly split into two sets each; $W_n = (y_1, \ldots, y_{r_n})$,
$Z_n = (x_1, \ldots, x_{r_n})$, $U_n = (y_{r_n+1}, \ldots, y_n)$, $V_n = (x_{r_n+1}, \ldots, x_n)$. Let
$r_n \to \infty$ and $r_n/n \to 0$, as $n \to \infty$. Estimate $J(u)$ using W_n and Z_n and
average the two estimates. Then determine the ω-vector used in the
rank test from the two remaining samples, U_n and V_n. Also, let
$[0 = h_{n,0} < h_{n,1} < \ldots < h_{n,q_n} = r_n]$ be a sequence of $q_n + 1$ - tuples of
integers, and let $y_{(\ )}$ be the order statistics of W_n. Finally, let
c_n be a sequence of constants and suppress the dependence of the
sample size on r_n and c_n by denoting them by r and c. Then the esti-
mate $J^H(u)$ is of the following form:

$$J^H(i/N-2r+1, W_n) =$$

$$(\tfrac{1}{2}) r^{-1/30} \{ 1/[y_{(h_{n,j+c})} - y_{(h_{n,j-c})}] - [y_{(h_{n,j+c+1})} - y_{(h_{n,j-c+1})}]\}$$

for $h_{n,j}/r < i/n-2r+1 < h_{n,j+1}/r$, $j = 2,\ldots,q_n$, $i = 1,\ldots,n-2r$

and $= 0$ otherwise .

The definition is completed by taking $J^H(u)$ to be constant on the intervals $[(i-1)/(n-2r), i/(n-2r)]$, $i=1,\ldots,n-2r$. Form a similar estimate based on $\{Z_n\}$ and average the two estimates. Van Eeden's procedure differs from this by monotonizing the function $J^H(u)$ and both Hajek and Van Eeden subtract the mean of the $J^H(i/n-2r+1)$'s from each estimated coefficient. The resulting two sample test can be written in the following form:

$$T_n(U,V) = \sum_{i=1}^{n-r} J^H(i/n-2r+1)\omega_{n,i}$$

where for $i=1,\ldots,n-K_n$, $\omega_{n,i}$ is the indicator of the i-th order statistic of the combined sample $\{U,V\}$.

We now study the differences between the estimator proposed here and the estimator developed by Hajek. There are three basic differences, two of which relate directly to the estimate of the score function, and a third (which will be discussed later) which deals with the construction of the R-estimate. To facilitate the comparison we will explicitly write the estimate proposed here as a function of the data. (We will omit for now the fact that both

9

estimates are averages of independent estimates from the two samples.) Letting $D_{k(n)}(a)$ be the distance from a to its $k(n)$-th nearest neighbor of $\underset{\sim}{X}$, we have

$$\hat{L}_n(i/n+1) = [\hat{f}(\hat{G}(\lambda_{j-1})) - \hat{f}(\hat{G}(\lambda_j))]/(\lambda_j - \lambda_{j-1})$$

$$= [\hat{f}(X_{(P(j-1))}) - \hat{f}(X_{(P(j))})]/(\lambda_j - \lambda_{j-1})$$

$$= [k(n)/2n \ D_{k(n)}(X_{(P(j-1))})$$

$$- k(n)/2n \ D_{k(n)}(X_{(P(j))})]/(\lambda_j - \lambda_{j-1})$$

$$= B([1/(X_{(P(j-1)+n(1))} - X_{(P(j-1)-n(2))})]$$

$$- [1/(X_{(P(j)+n(3))} - X_{(P(j)-n(4))})])$$

for $P(j-1) < i < P(j)$, where $n(1)$ and $n(2)$ (and similarly $n(3)$ and $n(4)$) denote the number of data points from $X_{(P(j-1))}$ to the left and to the right, which are passed in order to symmetrically enclose $k(n)$ data points. Here B is a constant depending only on the sample size. Note that with probability one, one of the end points of the interval

$$(X_{(P(j-1))-n(2))}, \ X_{(P(j-1)+n(1))})$$

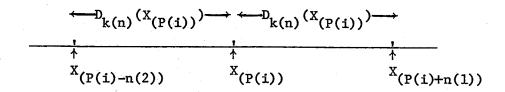is not a point of the data set. (See Figure 1.)

$$\longleftarrow D_{k(n)}(X_{(P(i))}) \longrightarrow \quad \longleftarrow D_{k(n)}(X_{(P(i))}) \longrightarrow$$

$$X_{(P(i)-n(2))} \qquad\qquad X_{(P(i))} \qquad\qquad X_{(P(i)+n(1))}$$

Figure 1

We recall that

$$J^H(i/N-2r+1, W_n)$$

$$= S\{1/[y_{(h_{n,j+c})} - y_{(h_{n,j-c})}] - 1/[y_{(h_{n,j+c+1})} - y_{(h_{n,j-c+1})}]\} \quad ,$$

for some constant S.

The similarities and differences between the proposed coefficients and the coefficients of Hajek's test can now be explained. Ignoring the leading constants, we concentrate on the four order statistics appearing in each of the estimates.

We have from Hajek that $h_{n,j+1} - h_{n,j}/r \sim r^{-1/6}$. Thus:

1) The distances between the points at which $J^H(u)$ changes approach zero asymptotically.

2) The number of ω_i that have the same coefficient in Hajek's test is asymptotically $(h_{n,j+1} - h_{n,j})/(1/n-2r+1)$ which approaches infinity.

We see that the Hajek test is similar to a grouped rank test with the number of groups slowly increasing to infinity, at the rate $r_n^{1/6}$. The following similarities and differences are now apparent:

11

1) In Hajek's test the score function is estimated at asymptotically infinitely many points and will therefore be asymptotically closer to $J(u)$ than $\hat{L}_n(u)$ for most values of u.

2) In the tails, where good estimation is vital for robustness, the Hajek interval $(X_{(h_{n,j-c})}, X_{(h_{n,j+c})})$ will have $X_{(h_{n,j})}$ close to one of the two end points, since it will take a longer interval to find c_n order statistics in one direction than in the other. The corresponding interval for the proposed estimate, i.e.,

$$(X_{(P(j)-n(2))}, X_{(P(j)+n(1))}) \quad ,$$

is symmetric about $X_{(P(j))}$. For small samples this should be an advantage for the proposed estimate.

3) The estimated coefficients of Hajek's test and the test proposed here both involve estimates of the derivative of the density at a point u. In both approaches this is accomplished by taking the difference of estimates of the density at the end points of an interval containing u. The interval used for the determination of the proposed coefficients is much wider than that used in Hajek's method. This may contribute to a more stable estimate for small sample sizes.

4) Finally, and most importantly, Hajek's estimate is based on r_n data points, where $r_n/n \to 0$. The estimate here is based on the total sample.

In summary, the two major differences are:

1) The proposed test sacrifices some asymptotic efficiency for small sample performance.

12

2) The proposed test uses the entire sample instead of a
vanishingly small fraction to estimate J(u) and should
therefore be more stable.

2.5. Convergence Properties

Now that the estimate has been defined, we need to examine
its convergence properties. We will show that the above adaptive
rank test coefficients converge to Gastwirth's asymptotically most
powerful group rank test (AMPGRT) coefficients for a wide class of
distributions.

We will need the following regularity assumptions:

1) F has a continuous density f > 0 for all x.

2) f is bounded.

3) f' exists and is bounded for all x.

4) $0 < \int (f'/f)^2 f dx = I(f) < \infty$.

Let $\hat{f}$ be any location equivariant density function estimate which
converges strongly, uniformly to f (e.g., the nearest neighbor
density estimate, Devroye and Wagner [6]). Let $X_1, \ldots, X_n$ be i.i.d.
~ F(x). Let $F^{-1} = G$ be defined as before.

Note: Since we are assuming f' exists and is bounded for all x, we
have as a by-product that f is Lipschitz of order one and therefore
absolutely continuous, and therefore uniformly continuous.

LEMMA 1. Under the additional assumption that $k(n) > \sqrt{n \log n}$ we
have

$$[\hat{L}_n(u) - L(u)] \to 0 \text{ w.p. } 1 \quad ,$$

where

$$L(u) = [f(G(\lambda_{i-1})) - f(G(\lambda_i))]/(\lambda_i - \lambda_{i-1}) \quad .$$

PROOF. In the following we will often use the elementary fact that if as $n \to \infty$

$$X_n \to X \text{ w.p. } 1$$

and

$$Y_n \to Y \text{ w.p. } 1 \quad ,$$

then

$$X_n + Y_n \to X + Y \text{ w.p. } 1 \quad .$$

It is well known that convergence w.p. 1 is preserved by continuous functions, that is, if $h(u)$ is continuous and $X_n \to X$ w.p. 1, then $h(X_n) \to h(X)$ w.p. 1. Devroye and Wagner [6] show that if the two conditions

$$(2.5.1) \qquad D_{k(n)}(u) \to 0 \text{ w.p. } 1 \quad \text{as} \quad n \to \infty \quad ,$$

where $D_{k(n)}$ is as defined in Section 2.4, and

$$(2.5.2) \qquad n(D_{k(n)}(u)^2)/\log n \to \infty \text{ w.p. } 1 \quad \text{as} \quad n \to \infty \quad ,$$

are satisfied, then

$$(2.5.3) \qquad \sup_u [\hat{f}(u) - f(u)] \to 0 \text{ w.p. } 1 \quad \text{as} \quad n \to \infty \quad .$$

14

From Theorem 1 of Moore and Yackel [21], from a result of Kiefer [18], we have that if $k(n)/\log \log n \to \infty$, then

$$(2.5.4) \qquad K(n)/(nH(D_{k(n)}(u))) \to 1 \text{ w.p. } 1 \quad ,$$

where

$$H(D_{k(n)}(u)) = \int_{S(D_{k(n)}(u))} f(x)\,dx \quad ,$$

where $S(D_{k(n)}(u)) = $ interval of length $2D_{k(n)}$ centered at u. From (2.5.4) we have that $H(D_{k(n)}(u)) \to 0$ w.p. 1. Since $f(x) > 0$ for all X, we clearly also have $D_{k(n)}(u) \to 0$ w.p. 1. From the same work we also have that

$$(2.5.5) \quad 2D_{k(n)}(u) \inf_{x \in S(D_{k(n)}(u))} f(x) < H(D_{k(n)}(u))$$

$$< 2D_{k(n)}(u) \sup_{x \in (D_{k(n)}(u))} f(x) \quad .$$

Also by (2.5.4) and (2.5.5) we have that $nD_{k(n)}(u)/k(n) \to 1$ w.p. 1. Therefore, $n^2 D_{k(n)}(u)^2/K(n)^2 \to 1$ w.p. 1. Thus to satisfy (2.5.2) above we need $n^2/k(n)^2$ to be $< n/\log n$. This is equivalent to $k(n) > \sqrt{n \log n}$. Therefore, we have shown that $\hat{f}(u)$ converges to $f(u)$ w.p. 1 uniformly.

Now consider $\hat{f}(\hat{G}(u)) - f(G(u))$. This is equal to $\hat{f}(\hat{G}(u)) - f(\hat{G}(u)) + f(\hat{G}(u)) - f(G(u))$. We will show that each of these two differences converge to 0 w.p. 1. The first difference converges to 0 w.p. 1 by the strong uniform convergence of nearest neighbor density estimators demonstrated above. The fact that $\hat{G}(u) - G(u)$

15

converges to 0 w.p. 1 is stated previously. Then since convergence

w.p. 1 is preserved by continuous functions, we have that

$f(\hat{G}(u)) - f(G(u))$ converges to 0 w.p. 1. Dividing by $\lambda_i - \lambda_{i-1}$ does

not change the convergence and the desired result follows. Q.E.D.

COROLLARY. $\hat{L}_n(u) \to L(u)$ in probability as $n \to \infty$.

2.6. Formation of the Test Statistic

Our final goal is to use the coefficients developed above, to

create a test statistic. This statistic will have the basic form

$$\sum_{i=1}^{N} \hat{L}_n(i/N+1)\omega_i \quad .$$

It is necessary to determine the distributional properties of this

random variable in order to use it in testing situations. The sta-

tistic may be used conditionally on the estimated coefficients to

produce a level α test for finite n. However, we are going to pro-

pose a slightly different procedure.

We would like to normalize the test statistic in order to

provide a large sample version of the statistic which can exploit the

approximate normality. That is, we wish to modify the statistic to

have mean zero and asymptotic variance one under the null hypothesis

of no change in location. Then in order to determine the level α

critical value for the test, we will generate a large number of repe-

titions of the modified statistic and determine its 1-α percentile

empirically. The following lemmas are helpful in allowing us to

substract the mean from the above statistic producing a statistic with mean 0.

LEMMA 2. If the fractiles $(\lambda_1,\ldots,\lambda_k)$ are symmetric about 1/2, that is, if $\lambda_i = 1 - \lambda_{k+1-i}$ and if f is symmetric about 0, that is, $f(x) = f(-x)$, and finally, if $c_i = c_{k+2-i}$, then

$$\sum_{i=1}^{k+1} c_i [f(G(\lambda_{i-1})) - f(G(\lambda_i))] = 0 \quad .$$

PROOF. Assume k is odd. Since f is symmetric, $G(u) = -G(1-u)$. Then we have

$$[f(G(\lambda_{i-1})) - f(G(\lambda_i))] = [f(G(1-\lambda_{k+1-(i-1)})) - f(G(1-\lambda_{k+1-i}))]$$

$$= [f(-G(\lambda_{k+1-(i-1)})) - f(-G(\lambda_{k+1-i}))]$$

$$= [f(G(\lambda_{k+1-(i-1)})) - f(G(\lambda_{k+1-i}))]$$

$$= [f(G(\lambda_{k+2-i})) - f(G(\lambda_{k+1-i}))] \quad .$$

The result follows since $c_i = c_{k+2-i}$. The case for k even is similar. Q.E.D.

Recalling that $N = 2n$, let

$$\overline{L} = \sum_{i=1}^{2n} L(i/N+1)/N = \sum_{i=1}^{k+1} L(\lambda_i)(P(i) - P(i-1) + 1)/N$$

and

$$(2.6.1) \quad \overline{L}_n = \sum_{i=1}^{2n} \hat{L}_n(i/N+1)/N = \sum_{i=1}^{k+1} \hat{L}_n(\lambda_i)(P(i) - P(i-1) + 1)/N \quad ,$$

17

where again $P(i) = [N\lambda_i] + 1$. We note that the results presented here have specified equal sample sizes for notational convenience. The derived results remain true for the more general situation under simple conditions, e.g., the bounding of the limiting proportion of each of the two samples away from zero.

LEMMA 3. Under the conditions of Lemma 1 and Lemma 2, $\overline{L}_n$ converges in probability to 0.

PROOF. We have

$$(2.6.2) \quad \overline{L}_n = \sum_{i=1}^{k+1} \hat{L}_n(\lambda_i)(P(i) - P(i-1) - 1)/N$$

$$= \sum_{i=1}^{k+1} \frac{[\hat{f}(\hat{G}(\lambda_{i-1})) - \hat{f}(\hat{G}(\lambda_i))]}{N(\lambda_i - \lambda_{i-1})} (P(i) - P(i-1) - 1) \quad .$$

We have

$$\frac{P_i - P_{i-1} - 1}{N(\lambda_i - \lambda_{i-1})} = \frac{[N\lambda_i] - [N\lambda_{i-1}] - 1}{N\lambda_i - N\lambda_{i-1}} \to 1 \quad \text{as} \quad n \to \infty \quad .$$

Also, we have shown that $\hat{f}(\hat{G}(\lambda_i))$ converges in probability to $f(G(\lambda_i))$. Therefore, $\overline{L}_n$ given by (2.6.2) converges in probability, as $n \to \infty$, to

$$(2.6.3) \quad \sum_{i=1}^{k+1} [f(G(\lambda_{i-1})) - f(G(\lambda_i))] \quad ,$$

which $= 0$ by Lemma 2 with $c_i = 1$ for all i.   Q.E.D.

Our test statistics now has the modified form

$$(2.6.4) \qquad T^1 = \sum_{i=1}^{N} (\hat{L}_n(i/N+1) - \overline{L}_n)\omega_i \quad .$$

Henceforth, $\hat{L}_N$ will indicate an estimate based on a simple average of estimates from each of the two samples of size n. Similarly, $\overline{L}_N$ will represent (2.6.1) with $\overline{L}_n$ replaced by $\overline{L}_N$. To show that the modified statistic has mean zero, we need one result.

LEMMA 4. ω_i and $\hat{L}_N(j/N+1)$ are pairwise independent, for all i,j, when there is no difference in location in the two samples.

PROOF. We have

$$\text{Prob}\{\hat{L}_N(u)\epsilon(a,b); \ \omega_i = 1\} + \text{Prob}\{\hat{L}_N(u)\epsilon(a,b); \ \omega_i = 0\}$$

$$= \text{Prob}\{\hat{L}_N(u)\epsilon(a,b)\} \quad .$$

We also have under the null hypothesis that

$$\text{Prob}\{\hat{L}_N(u)\epsilon(a,b); \ \omega_i = 1\} = \text{Prob}\{\hat{L}_N(u)\epsilon(a,b); \ \omega_i = 0\} \quad .$$

Therefore,

$$\text{Prob}\{\hat{L}_N(u)\epsilon(a,b); \ \omega_i = 1\} = \frac{1}{2}\text{Prob}\{\hat{L}_N(u)\epsilon(a,b)\}$$

$$= \text{Prob}\{\omega_i = 1\} \cdot \text{Prob}\{\hat{L}_N(u)\epsilon(a,b)\} \quad \text{Q.E.D.}$$

We now show that the test statistic given by (2.6.4) does indeed have mean 0.

LEMMA 5. Let

$$\hat{T}_N = \sum_{i=1}^{N} (\hat{L}_N(i/N+1) - \overline{L}_N)\omega_i \quad .$$

Then, under the conditions of Lemma 1 and Lemma 2, if there is no difference in location between the samples, $E(\hat{T}_N) = 0$.

PROOF. By the independence shown in Lemma 4,

$$E(\hat{T}_N) = \sum_{i=1}^{N} E\{(\hat{L}_N(i/N+1) - \overline{L}_N)\omega_i\}$$

$$= \sum_{i=1}^{N} E\{(\hat{L}_N(i/N+1) - \overline{L}_N)E(\omega_i)\}$$

$$= (1/2) \sum_{i=1}^{N} E\{(\hat{L}_N(i/N+1) - \overline{L}_N)\}$$

$$= 0 \quad . \quad \text{QED.}$$

Now that we can assure ourselves of a statistic that has mean 0, we would like to modify the test statistic so that is has asymptotic variance 1. To do this we determine the asymptotic standard deviation and divide the test statistic by an estimate of this constant. We will, in fact, show that the asymptotic variance is

$$(\tfrac{1}{4}) \sum_{i=1}^{k+1} [f(G(\lambda_{i-1})) - f(G(\lambda_i))]^2/(\lambda_i - \lambda_{i-1}) \quad .$$

Therefore, if we can find a function of the data which converges to the above variance, we can divide by the square root of that function

of the data so that the resulting test statistic will have asymptotic variance 1.

LEMMA 6.   Under the conditions of Lemma 1 and Lemma 2,

$$\sum_{i=1}^{N} [\hat{L}_N(i/N+1) - \bar{L}_N]^2/N$$

converges in probability to

$$\sum_{i=1}^{k+1} [f(G(\lambda_{i-1})) - f(G(\lambda_i))]^2/(\lambda_i - \lambda_{i-1})\quad .$$

PROOF.

$$\sum_{i=1}^{N} [\hat{L}_N(i/N+1) - \bar{L}_N]^2/N$$

$$= \{\sum_{i=1}^{N} [\hat{L}_N(i/N+1)]^2 - 2\bar{L}_N \sum_{i=1}^{N} (\hat{L}_N(i/N+1))^2 + \sum_{i=1}^{N} (\bar{L}_N)^2/N$$

$$= \sum_{i=1}^{N} [\hat{L}_N(i/N+1)]^2/N - \bar{L}_N^2$$

We have shown that $\bar{L}_N$ converges in probability to 0. Also, since x^2 is a continuous function, and continuity preserves convergence in probability, we have that $\bar{L}_N^2$ converges to 0 in probability. Therefore, we can concentrate on the quantity

$(2.6.5) \quad \sum_{i=1}^{N} \hat{L}_N(i/N+1)^2/N$

$$= \sum_{j=1}^{k+1} \{(\hat{L}_n(\lambda_j, \underset{\sim}{X}) + \hat{L}_n(\lambda_j, \underset{\sim}{Y}))/2\}^2 (P(j) - P(j-1) - 1)/N$$

$$= \sum_{j=1}^{k+1} \{(\hat{L}_n(\lambda_j, \underset{\sim}{X}) + \hat{L}_n(\lambda_j, \underset{\sim}{Y}))/2\}^2 ([N\lambda_j] - [N\lambda_{j-1}] - 1)/N \quad,$$

where $\hat{L}_n(\lambda_j, \underset{\sim}{X})$ is written to explicitly indicate the dependence on the first sample, and similarly for $\hat{L}_n(\lambda_j, \underset{\sim}{Y})$. We have that $\hat{L}_n(\lambda_j, \underset{\sim}{X})^2$ and $\hat{L}_n(\lambda_j, \underset{\sim}{Y})^2$ both converge in probability to $L(\lambda_j)^2$ and therefore $\Sigma_{j=1}^{j+1} \hat{L}_N(\lambda_j)^2$ converges in probability to $\Sigma_{j=1}^{k+1} L(\lambda_j)^2$. For large N we can replace $([N\lambda_j] - [N\lambda_{j-1}] - 1)/N$ by $\lambda_j - \lambda_{j-1}$ so that expression (2.6.5) is asymptotically equivalent to $\Sigma_{j=1}^{k+1} \hat{L}_N(\lambda_j^2)(\lambda_j - \lambda_{j-1})$. Hence, as $n \to \infty$, $\Sigma_{i=1}^{N} \hat{L}_N(i/N+1)^2/N$ converges in probability to $\Sigma_{j=1}^{k+1} L(\lambda_j)^2 (\lambda_j - \lambda_{j-1})$ which equals $\Sigma_{j=1}^{k+1} [f(G(\lambda_{j-1}) - f(G(\lambda_j)]^2/(\lambda_j - \lambda_{j-1})$. Q.E.D.

Thus a studentized version of the original test statistic is

$$T^2 = \frac{\sum\limits_{i=1}^{N} [\hat{L}_N(i/N+1) - \bar{L}_N]\omega_i/N^{\frac{1}{2}}}{(\frac{1}{4})\{(\sum\limits_{i=1}^{N} [\hat{L}_N(i/N+1) - \bar{L}_N]^2)/N\}^{\frac{1}{2}}} \quad .$$

We have shown that the numerator of T^2 has mean zero. We will show later that T^2 has asymptotic variance = 1.

2.7. Location and Scale Equivariance of the Test Statistic

Consider the test statistic

$$\tilde{T}_N = \frac{\sum\limits_{i=1}^{N} \hat{L}_N(i/N+1)\omega_i}{[1/4) \sum\limits_{i=1}^{N} \hat{L}_N(i/N+1)^2]^{\frac{1}{2}}}$$

where

$$\hat{L}_N(i/N+1) = \frac{[\hat{L}_n(i/N+1,\underset{\sim}{X}) + \hat{L}_n(i/N+1,\underset{\sim}{Y})]}{2}$$

and

$$\hat{L}_n(i/N+1,\underset{\sim}{X}) = [\hat{f}(\hat{G}(\lambda_{j-1})) - \hat{f}(\hat{G}(\lambda_j))]/(\lambda_j - \lambda_{j-1})$$

for $\lambda_{j-1} < i/N+1 \leq \lambda_j$ (similarly for $\hat{L}_n(i/N+1,\underset{\sim}{Y})$. We have $\hat{G}_{c\underset{\sim}{X}+d}(u) = c\hat{G}_{\underset{\sim}{X}}(u) + d$ if $c > 0$. Similarly with $\underset{\sim}{Y}$ replacing $\underset{\sim}{X}$. Also, $\hat{f}_{c\underset{\sim}{X}+d}(cu+d) = \hat{f}_{c\underset{\sim}{X}}(cu) = k(N)/cND_{k(N)}(u) = (1/c)\hat{f}_{\underset{\sim}{X}}(u)$. Similarly with $\underset{\sim}{Y}$ replacing $\underset{\sim}{X}$. Thus,

$$\hat{L}_n(u,c\underset{\sim}{X}+d) = \frac{\hat{f}_{c\underset{\sim}{X}+d}(\hat{G}_{c\underset{\sim}{X}+d}(\lambda_{i-1})) - \hat{f}_{c\underset{\sim}{X}+d}(\hat{G}_{c\underset{\sim}{X}+d}(\lambda_i))}{(\lambda_i - \lambda_{i-1})}$$

for $\lambda_{i-1} < u \leq \lambda_i$

$$= \frac{\hat{f}_{c\underset{\sim}{X}+d}(c\hat{G}_{\underset{\sim}{X}}(\lambda_{i-1})+d) - \hat{f}_{c\underset{\sim}{X}+d}(c\hat{G}_{\underset{\sim}{X}}(\lambda_i)+d)}{(\lambda_i - \lambda_{i-1})}$$

$$= \frac{(1/c)\hat{f}_{\underset{\sim}{X}}(\hat{G}_{\underset{\sim}{X}}(\lambda_{i-1})) - (1/c)\hat{f}_{\underset{\sim}{X}}(\hat{G}_{\underset{\sim}{X}}(\lambda_i))}{(\lambda_i - \lambda_{i-1})}$$

$$= (1/c)\hat{L}_n(u,\underset{\sim}{X}) \quad .$$

Similarly we have $\hat{L}_n(u,c\underset{\sim}{Y}+d) = (1/c)\hat{L}_n(u,\underset{\sim}{Y})$. Clearly,

$\omega_i(c\underset{\sim}{X}+d,c\underset{\sim}{Y}+d) = \omega_i(\underset{\sim}{X},\underset{\sim}{Y})$ for $c > 0$. The denominator

$[(1/4) \Sigma_{i=1}^{N} \hat{L}_N(i/N+1)^2]^{\frac{1}{2}}$ also clearly becomes multiplied by $1/c$.

Therefore, for $c > 0$, the test statistic will not change. If we were

to change $\hat{L}_N(i/N+1)$ by subtracting its mean $\bar{L}_N$, all the equivariance

properties shown above would remain since the average of equivariant

terms is also equivariant. Equivariance under a location shift will

be used later to help demonstrate the asymptotic properties of the

location estimate resulting from $\tilde{T}_N$.

3. ASYMPTOTIC PROPERTIES OF PROPOSED RANK TEST

3.1. Asymptotic Equivalence of the Test Statistic and the Asymptotically Most Powerful Grouped Rank Test Statistic

Let us assume that $\{\underset{\sim}{X}\} = \{X_1,\ldots,X_n\}$, $\{\underset{\sim}{Y}\} = \{Y_1,\ldots,Y_n\}$, are two samples of i.i.d. observations from the c.d.f. F. Let F satisfy the following regularity conditions:

1) F has a continuous density f > 0 for all X.

2) f' is bounded for all x.

3) $0 < \int (f'/f)^2 f\,dx = I(f) < \inf$.

4) f is symmetric about 0.

5) Finally, assume $\lambda_i = 1 - \lambda_{k+1-i}$, $i = 1,2,\ldots,k$.

Let $\hat{G}(u) = \hat{F}^{-1}(u) = X_{([nu]+1)}$, and let $\hat{f}(X)$ be the k(N) nearest neighbor density function estimator. Let $\hat{L}_N(u) = (\hat{L}_n(u,\underset{\sim}{X}) + \hat{L}_n(u,\underset{\sim}{Y}))/2$ where $\hat{L}_n(u,\underset{\sim}{X}) = [\hat{f}(\hat{G}(\lambda_{i-1})) - \hat{f}(\hat{G}(\lambda_i))]/(\lambda_i - \lambda_{i-1})$ for $\lambda_{i-1} < u < \lambda_i$ and similarly for $\hat{L}_n(u,\underset{\sim}{Y})$. Let

$$(3.1.1) \qquad L^*(u) = \hat{L}_N(u) - \overline{L}_N \quad ,$$

where $\overline{L}_N$ is given by (2.6.1) with $\hat{L}_n$ replaced by $\hat{L}_N$

Note: The number of fractiles k should not be confused with k(N), the number of nearest neighbors used.

Let

$$T^* = \sum_{i=1}^{N} L^*(i/N+1)\omega_i / N^{\frac{1}{2}}$$

and

$$S = \sum_{i=1}^{N} L(i/N+1)\omega_i/N^{\frac{1}{2}} \quad .$$

Thus T^* is the proposed test and S is the AMPGRT.

THEOREM 1. Assuming conditions 1) through 5) above hold, we have that $T^* - S \to 0$ in probability.

PROOF. We have

$$(3.1.2) \qquad T^* - S = \sum_{i=1}^{k+1} \frac{L^*(\lambda_i) - L(\lambda_i)}{N^{\frac{1}{2}}} \left[\sum_{j=P(i-1)+1}^{P(i)} \omega_j \right] \quad .$$

We first show that the mean of (3.1.2) is asymptotically 0 so that we can subtract its mean without changing its convergence properties. This is equivalent to subtracting .5 from each of the ω_i's in equation (3.1.2). Thus we would like to show that

$$\sum_{i=1}^{k+1} \frac{L^*(\lambda_i)}{N^{\frac{1}{2}}} [\sum_{j=P(i-1)+1}^{P(i)} (1/2)] - \sum_{i=1}^{k+1} \frac{L(\lambda_i)}{N^{\frac{1}{2}}} [\sum_{j=P(i-1)+1}^{P(i)} (1/2)]$$ is

asymptotically zero. By definition (3.1.1)

$$(3.1.3) \qquad \sum_{i=1}^{k+1} L^*(\lambda_i)(P(i) - P(i-1) - 1)/2N^{\frac{1}{2}} = 0 \quad ,$$

since we have subtracted the mean from the coefficients. Therefore, it remains to show that $A_n = \sum_{i=1}^{k+1} \frac{L(\lambda_i)}{N^{\frac{1}{2}}} [\sum_{j=P(i-1)+1}^{P(i)} (1/2)]$ is asymptotically zero.

We have shown by Lemma 2 that

$$\sum_{i=1}^{k+1} c_i [f(G(\lambda_{i-1})) - f(G(\lambda_i))] = 0 \quad ,$$

when $c_i = c_{k+2-i}$, for all i. Since $\lambda_i = 1 - \lambda_{k+1-i}$ we have for each i

$$(N\lambda_i - N\lambda_{i-1} - 1)/(\lambda_i - \lambda_{i-1}) = \frac{N(1-\lambda_{k+1-i}) - N(1-\lambda_{k+2-i}) - 1}{(\lambda_{k+2-i} - \lambda_{k+1-i})} \quad .$$

Therefore the sequence of coefficients

$$c_i = (N\lambda_i - N\lambda_{i-1})/2(\lambda_i - \lambda_{i-1})$$

possesses the property that $c_i = c_{k+2-i}$ for all i and hence by Lemma 2,

$$\sum_{i=1}^{k+1} [(N\lambda_i - N\lambda_{i-1} - 1)/2(\lambda_i - \lambda_{i-1})][f(G(\lambda_{i-1})) - f(G(\lambda_i))] = 0 \quad .$$

The above expression may be rewritten as $\sum_{i=1}^{k+1} L(\lambda_i)(N\lambda_i - N\lambda_{i-1} - 1)/2N^{\frac{1}{2}}$, which is asymptotically equal to A_n. Therefore,

$\sum_{i=1}^{k+1} \dfrac{[L^*(\lambda_i) - L(\lambda_i)]}{N^{\frac{1}{2}}} (P(i) - P(i-1) - 1)/2$ converges to 0 as $n \to \infty$.

Subtracting this expression from (3.1.2) we obtain

$$(3.1.4) \qquad \sum_{i=1}^{k+1} \frac{L^*(\lambda_i) - L(\lambda_i)}{N^{\frac{1}{2}}} \left[\sum_{j=P(i-1)+1}^{P(i)} \omega_i \right]$$

$$- \sum_{i=1}^{k+1} \frac{[L^*(\lambda_i) - L(\lambda_i)]}{N^{\frac{1}{2}}} \frac{(P(i) - P(i-1) - 1)}{2}$$

$$= \sum_{i=1}^{k+1} [L^*(\lambda_i) - L(\lambda_i)] \left[\sum_{j=P(i-1)+1}^{P(i)} (\omega_i - 1/2) \right]/N^{\frac{1}{2}} \quad .$$

27

We have shown that the difference of (3.1.4) and (3.1.2) converges to 0 in probability. If we can show that

$$(3.1.5) \qquad \sum_{j=P(i-1)+1}^{P(i)} (\omega_j - 1/2)/N^{\frac{1}{2}}$$

is bounded in probability, then since $L^*(u)$ converges in probability to $L(u)$ we have that (3.1.4) is the sum of $k+1$ random variables which converge to 0 in probability, and hence itself converges to 0 in probability. However, it is immediate that (3.1.5) is bounded in probability since it is clearly a linear rank test statistic. Therefore, by Theorem V.1.6. of Hajek and Sidak we have that (3.1.5) has bounded variance. Hence by Chebychev's inequality, (3.1.5) must be bounded in probability. Thus we have shown that $T^* - S$ converges to 0 in probability for the wide class of distributions satisfying conditions 1) through 5). Q.E.D.

Now we anticipate a result needed to utilize results of Hodges and Lehmann [12]. For this purpose we introduce the following framework: We consider the two samples $\underset{\sim}{X}$ and $\underset{\sim}{Y}$; but instead of a zero value for the shift, d, of the Y's relative to the X's, we let d depend on the total sample size N according to $d_N = 2b/[N \, I(f)]^{\frac{1}{2}}$. In order to use results of Hodges and Lehmann [12], it is necessary to demonstrate that the conclusion of Theorem 1 remains valid under this model. That is, we require the following:

COROLLARY. Assume that we have the sequence of alternatives $d_N = 2b/[N \, I(f)]^{\frac{1}{2}}$ as described above. Then under assumptions 1) through 5) we have that $T^* - S$ converges to 0 in probability as $N \to \infty$.

PROOF. By Theorem VI.2.3, application #1, of Hajek and Sidak [11],
we have that (3.1.5) converges to a distribution with finite
variance. Therefore, since Lemma 2 is independent of the difference
in location between the two samples, the argument in the proof of
Theorem 1 remains valid and we have that under the sequence of alter-
natives above, $T^* - S$ converges to 0 in probability.   Q.E.D.

To study the asymptotic distribution of the proposed test T^*
we can therefore concentrate on the asymptotic distribution of the
AMPGRT, i.e.,

$$S = \sum_{i=1}^{k+1} L(\lambda_i) \left[\sum_{j=P(i-1)+1}^{P(i)} \omega_j \right] / N^{\frac{1}{2}} \quad ,$$

since we have now shown that the difference between T^* and S con-
verges to 0 in probability.

THEOREM 2. If conditions 1) through 5) hold, then T^* has an asymp-
totic normal distribution with mean 0 and variance

$$(1/4) \int L^2(u) du \quad .$$

PROOF. Clearly we have that S has the same asymptotic distribution
as

$$(3.1.6) \quad S^1 = \sum_{i=1}^{k+1} L(\lambda_i) \frac{N\lambda_i - N\lambda_{i-1} - 1}{P(i) - P(i-1) - 1} \left[\sum_{j=P(i-1)+1}^{P(i)} \omega_j \right] / N^{\frac{1}{2}} \quad ,$$

since the middle factor converges to 1 as N goes to infinity. We
also have that $E(S^1) = 0$ since

$$\sum_{j=P(i-1)+1}^{P(i)} E(\omega_j) = \sum_{j=P(i-1)+1}^{P(i)} (1/2) = (P(i) - P(i-1) - 1)/2$$

so that

$$E(S^1) = \sum_{i=1}^{k+1} L(\lambda_i)(N\lambda_i - N\lambda_{i-1})/2$$

$$= \sum_{i=1}^{k+1} [f(G(\lambda_{i-1})) - f(G(\lambda_i))]N/2$$

$$= 0 \quad \text{by Lemma 2 where} \quad c_i = N/2 \quad .$$

It is clear that $L(u)$ is square integrable since it is a finite step function on a bounded interval. Let

$$V_N(R_{N,i}) = [f(G(\lambda_{j-1})) - f(G(\lambda_j))]/(\lambda_j - \lambda_{j-1})$$

for $P(j-1) < R_{N,i} < P(j)$. Then $V_N(1 + [uN])$ and $L(u)$ are the same step function except that the change in steps occurs at points whose differences are converging to 0. Therefore we clearly have that

$$\lim_{N \to \infty} \int_0^1 [V_N(1 + [uN])] - L(u)]^2 \, du = 0 \quad .$$

This satisfies the conditions for Theorem V.1.6. in Hajek and Sidak [11] with their

$$c_i = \begin{cases} 1 & \text{for} \quad i = 1, \ldots, N/2 \\ 0 & \text{for} \quad i = (N/2) + 1, \ldots, N \quad . \end{cases}$$

Therefore, we have from this result that S has an asymptotic normal distribution with expectation 0 and variance

$$(1/4) \int_0^1 L(u)^2 \, du = (1/4) \sum_{i=1}^{k+1} [f(G(\lambda_{i-1})) - f(G(\lambda_i))]^2 / (\lambda_i - \lambda_{i-1}) \quad .$$

Q.E.D.

As before, it is necessary to establish the corresponding results for the sequence of non-zero alternatives d_N. To this end we consider the following:

COROLLARY. Given the sequence of alternatives $d_N = 2b/[N \, I(f)]^{\frac{1}{2}}$, then if conditions 1) through 5) are satisfied, we have that T^* has an asymptotic normal distribution with

$$\text{mean} = b/[2 \, I(f)^{\frac{1}{2}}] \int_0^1 \left[-L(u) \frac{f'(G(u))}{f(G(u))} \right] du \quad ,$$

and

$$\text{variance} = (1/4) \int_0^1 L^2(u) \quad .$$

PROOF. By Theorem VI.2.3., application #1, of Hajek and Sidak [11], the desired result follows.

3.2. Asymptotic Relative Efficiency

We will now demonstrate that as the number of fractiles is increased to infinity, the efficiency of

$$(3.2.1) \qquad T^* = \sum_{i=1}^{N} L^*(i/N+1)\omega_i / N^{\frac{1}{2}}$$

compared to the LMPRT converges to 1.

THEOREM 3. Under conditions 1) through 5) of Section 3.1, as

$\max[\lambda_i - \lambda_{i-1}]$ goes to 0, the asymptotic relative efficiency of the

adaptive test T^* given by (3.2.1) approaches 1.

PROOF. We have shown that

$$T^* = \sum_{i=1}^{k+1} L^*(\lambda_i) \left(\sum_{j=P(i-1)}^{P(i)} \omega_i \right) / N^{\frac{1}{2}}$$

has the same asymptotic distribution properties under the null hypo-

theses and under the sequence of alternatives d_N given by (3.1.9) as

the AMPGRT

$$S = \sum_{i=k}^{N} L(i/N+1)\omega_i / N^{\frac{1}{2}} \quad .$$

The asymptotic relative efficiency of $\hat{T}$ is therefore the same as that

of S, which, by Gastwirth [8], is

$$I(f)^{-1} \sum_{i=1}^{k(m)+1} [f(G(\lambda_{i-1}(m))) - f(G(\lambda_i(m)))]^2 / (\lambda_i(m) - \lambda_{i-1}(m)) \quad .$$

Now assume that we have sequences $\{\lambda_i(m)\}$ where

$m = \max\{[\lambda_i(m) - \lambda_{i-1}(m)]\}$. The number of elements of each partition

is then $k(m) + 1$. We want to show that

$$I(f)^{-1} \sum_{i=1}^{k(m)+1} [f(G(\lambda(m)_{i-1})) - f(G(\lambda(m)_i))]^2 / (\lambda(m)_i - \lambda(m)_{i-1}) \to 1$$

$$\text{as } m \to 0 \quad .$$

Since we have the existence of f', we know from the mean value

theorem that

32

$$f(G(\lambda_i(m))) - f(G(\lambda_{i-1}(m))) = [\lambda_i(m) - \lambda_{i-1}(m)]f'(G(U_i))/f(G(U_i))$$

where $\lambda_{i-1}(m) < U_i < \lambda_i(m)$. Therefore,

$$\sum_{i=1}^{k(m)+1} [f(G(\lambda_{i-1}(m))) - f(G(\lambda_i(m)))]^2/(\lambda_i(m) - \lambda_{i-1}(m))$$

$$= \sum_{i=1}^{k(m)+1} \{[\lambda_i(m) - \lambda_{i-1}(m)]^2/[\lambda_i(m) - \lambda_{i-1}(m)]\}$$

$$\cdot [f'(G(U_i))/f(G(U_i))]^2$$

$$= \sum_{i=1}^{k(m)+1} [\lambda_i(m) - \lambda_{i-1}(m)][f'(G(U_i))/f(G(U_i))]^2 \quad,$$

which is the Riemann approximating sum for $I(f)$ as m goes to 0
Q.E.D.

It is important to point out that even though the A.R.E. $\rightarrow$ 1, the rate is a function of the convergence of the Riemann sum to the limiting integral, which will depend on the smoothness of $f'(G(u))/f(G(u))$. However, given a finite set of distributions, e.g., normal, Cauchy, logistic, and Laplace, we can guarantee that we will be $\in$ away from 1 in A.R.E. for all the distributions under consideration, for any $\in > 0$, by choosing the finest partition needed for any member in that set.

We are sacrificing some degree of asymptotic optimality in order to achieve better small sample performance. We note that W. Albers [1], in a recent article in the Annals of Statistics believes that this is inevitable for applied work in this area.

Actually, the score functions of the distributions that are typically

used in robustness investigations are very smooth functions and in an

important sense we are not giving up much in terms of asymptotic

optimality. As we see below, the asymptotic relative efficiencies,

for k very small, are quite close to 1. We conjecture that for dis-

tributions that satisfy conditions 1) through 5) of Section 3.1 there

exists a sequence of fractiles depending on N that will enable the

procedure described here to have full asymptotic efficiency.

J. W. Tukey in a private communication mentioned that spacings

between fractiles of the order of $N^{-1/3}$ might work.

In order to determine how much efficiency is being sacrificed

it is necessary to first compute the asymptotically most powerful

grouped rank tests explicitly for these distributions of interest.

Here we will only consider k odd and $\lambda_{k+1/2} = 1/2$ and concentrate on

the four distributions mentioned above.

For the Laplace distribution, since the score function of the

asymptotically most powerful rank test is $\text{sign}(u - 1/2)$, we have that

the LMPRT is also the AMPGRT, since all the ω_i for $0 < i/N + 1 < 1/2$

have the same coefficient, and similarly for ω_i with $1/2 < i/N + 1 < 1$.

Therefore, the A.R.E. for all k is equal to 1. For the three other

distributions more work is needed. For the logistic distribution we

have

$$[f(G(\lambda_{i-1})) - f(G(\lambda_i))]/(\lambda_i - \lambda_{i-1})$$

$$= [\lambda_{i-1}(1 - \lambda_{i-1}) - \lambda_i(1 - \lambda_i)](\lambda_i - \lambda_{i-1}) \quad .$$

For the Cauchy distribution we have

$$[f(G(\lambda_{i-1})) - f(G(\lambda_i))]/(\lambda_i - \lambda_{i-1})$$

$$= 1/\pi \ [\cos^2(\pi(\lambda_{i-1} - 1/2)) - \cos^2(\pi(\lambda_i - 1/2))]/(\lambda_i - \lambda_{i-1}) \quad .$$

For the normal distribution we have

$$[f(G(\lambda_{i-1})) - f(G(\lambda_i))]/(\lambda_i - \lambda_{i-1})$$

$$= (1/2 \ \pi)^{\frac{1}{2}} \ [\exp(-1/2(\Phi^{-1}(\lambda_{i-1}))^2) - \exp(-1/2(\Phi^{-1}(\lambda_i))^2)]/(\lambda_i - \lambda_{i-1})$$

where Φ^{-1} represents the inverse normal distribution.

For k = 3 we have for regularly spaced λ_i:

$$\lambda_0 = 0, \ \lambda_1 = .25, \ \lambda_2 = .50, \ \lambda_3 = .75, \ \lambda_4 = 1.00 \quad .$$

For these λ_i the coefficients for the logistic AMPGRT are

$$L(.25) = -.75, \ L(.5) = -.25, \ L(.75) = .25, \ L(1) = .75 \quad .$$

The asymptotic variance is .3125. The efficiency is .9375.

The coefficients for the Cauchy AMPGRT are

$$L(.25 = -.636, \ L(.5) = -.636, \ L(.75) = .636, \ L(1) = .636 \quad .$$

The asymptotic variance is .404. The efficiency is .8080.

The coefficients for the normal AMPGRT are

$$L(.25) = -1.27, \ L(.5) = -.325, \ L(.75) = .325, \ L(1) = 1.27 \quad .$$

The asymptotic variance is .8593. The efficiency is also .8593.

For k = 7 we have for regularly spaced λ_i;

$$\lambda_0 = 0, \ \lambda_1 = .125, \ \lambda_2 = .250, \ \lambda_3 = .375, \ \lambda_4 = .50,$$

$$\lambda_5 = .625, \ \lambda_6 = .750, \ \lambda_7 = .875, \ \lambda_8 = 1.00 \ .$$

The coefficients for the logistic AMPGRT are

$$L(.125) = -.875, \ L(.25) = -.625, \ L(.375) = -.375, \ L(.500) = -.125,$$

$$L(.625) = .125, \ L(.75) = .375, \ L(.875) = .625, \ L(1.00) = .875.$$

The asymptotic variance is .3282. The efficiency is .9846.

The coefficients for the Cauchy AMPGRT are

$$L(.125) = -.373, \ L(.25) = -.899, \ L(.375) = -.902, \ L(.500) = -.373,$$

$$L(.625) = .373, \ L(.75) = .902, \ L(.875) = .899, \ L(1.00) = .373.$$

The asymptotic variance is .475. The efficiency is .9500.

The coefficients for the normal AMPGRT are

$$L(.125) = -1.64, \ L(.25) = -.889, \ L(.375) = -.496, \ L(.500) = -.159,$$

$$L(.625) = .159, \ L(.75) = .496, \ L(.875) = .889, \ L(1.00) = 1.64.$$

The asymptotic variance is .9378. The efficiency is also .9378.

Below the proposed test is denoted by PT and its power is compared to that for the LMPRT on the distribution for which that test is locally most powerful, for various distributions. The first half of each section below relates to the determination of the 95% point under the null hypothesis for each of the two tests for each distribution, in order to guarantee that the tests have the same

36

size. The second half of each section gives the empirical power for the two tests. The ratio of the power is also given.

3.3. Monte Carlo Investigation of Adaptive Testing

In the following table $\hat{F}^{-1}$ denotes the percentile for that particular distribution under study. PT denotes the proposed test.

TABLE 1

RESULTS OF SIMULATIONS OF RANK TESTS

Fractiles = {.00, .25, .50, .75, 1.0}

Sample Size = 20 For Each Sample

Normal $\dfrac{1}{(2\pi)^{\frac{1}{2}}} \exp(-(1/2)(x-d)^2)$ LMPRT = Normal Scores Test

15,000 repetitions   d = 0   → $\hat{F}^{-1}$ (.95) of LMPRT = 1.62594

                                    $\hat{F}^{-1}$ (.95) of PT   = 1.81353

15,000 repetitions   d = .40   Pr{LMPRT > 1.62594} = .3455

                                    Pr{PT   > 1.81353} = .2936

                                    Ratio = .8497

Cauchy $\dfrac{1}{\pi} (1/(1+(x-d)^2))$

15,000 repetitions   d = 0   → $\hat{F}^{-1}$ (.95) of LMPRT = 1.65900

                                    $\hat{F}^{-1}$ (.95) of PT   = 1.78258

15,000 repetitions   d = .75   Pr{LMPRT > 1.64900} = .4481

                                    Pr{PT   > 1.78258} = .3845

                                    Ratio = .8582

Laplace $\dfrac{1}{2} \exp(-[x-d])$ LMPRT = Median Test

15,000 repetitions   d = 0   → F^{-1} (.95) of LMPRT = 1.700 (exact)

                                    F^{-1} (.95) of PT   = 1.74412

15,000 repetitions   d = .75   Pr{LMPRT > 1.62594} = .6484

                                    Pr{PT   > 1.81353} = .6135

                                    Ratio = .9462

Logistic $\dfrac{\exp(-(x-d))}{(1+\exp(-(x-d)))^2}$ LMPRT = Wilcoxon Two-Sample Test

15,000 repetitions   d = 0   → $\hat{F}^{-1}$ (.95) of LMPRT = 1.66006

                                    $\hat{F}^{-1}$ (.95) of PT   = 1.79796

15,000 repetitions   d = .40   →   Pr{LMPRT > 1.62594} = .3694

                                    Pr{PT   > 1.81353} = .3402

                                    Ratio = .9210

4. FORMATION OF ADAPTIVE R-ESTIMATE
AND RESULTING PROPERTIES

4.1. The Derived Two-Sample Location Estimate

Now that we have constructed a two-sample test for change in location which has some adaptive properties, we investigate the transfer of these properties to the corresponding estimation problem. To illustrate the connection between non-parametric testing and estimation, consider the Wilcoxon two-sample rank sum test statistic which can be written as

$$W(\underset{\sim}{X},\underset{\sim}{Y}) = \sum_{i=1}^{N} (i/N+1)\omega_i$$

where $N = 2n$ and n is the number of $\underset{\sim}{X}$'s and $\underset{\sim}{Y}$'s. Now let $\omega_i(e)$ denote the vector of indicator functions which would result from a shift of e in the second sample so that

$$W(\underset{\sim}{X},\underset{\sim}{Y}+e) = \sum_{i=1}^{N} (i/N+1)\omega_i(e) \quad .$$

Let $w_N^* = \inf\{e : W(\underset{\sim}{X},\underset{\sim}{Y}+e) > 0\}$, $w_N^{**} = \sup\{e : W(\underset{\sim}{X},\underset{\sim}{Y}+e) < 0\}$. Then $w_N = (w_N^* + w_N^{**})/2$ is known as the Hodges-Lehmann estimator [12] in the two-sample case. This estimator should be thought of as that shift in the second sample which causes the test statistic for change in location to assume its smallest absolute value. This procedure can, of course, be used in almost any location testing framework in order to derive a point estimate of location from a test statistic. In particular, we can use the above procedure for the two-sample test

proposed here. A point estimate of location based on a rank test will be called an R-estimate. An analogous procedure may be used for the one-sample case. For this case, in order to use the theory outlined above, which makes particular use of the two-sample framework, two samples must be artificially created from the single sample. We will concentrate on the method for accomplishing this below.

Suppose we have the sample $\underset{\sim}{X} = x_1, x_2, \ldots, x_n$. The second sample then consists of the negative values of the first sample. Then $\omega_i(e)$ is the indicator of whether the i-th order statistic of the combined sample of $\{x_1 - e, \ x_2 - e, \ldots, x_n - e\}$ and $\{-x_1 + e, \ -x_2 + e, \ldots, -x_n + e\}$ comes from the original sample or the second artificial sample. This method is more fully discussed in Huber [16], where the influence curve and breakdown points for R-estimates constructed in this manner are given.

We now proceed with the asymptotic theory of the proposed R-estimate. Let

$$(4.1.1) \qquad \hat{T}(\underset{\sim}{X}, \underset{\sim}{Y}) = \sum_{i=1}^{k+1} L^*(\lambda_i) \left\{ \sum_{j=P(i-1)+1}^{P(i)} \omega_j \right\} ,$$

where $L^*(\lambda_i)$ previously defined by (3.1.1) is again an average of estimates from the $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ samples. We denote this test statistic by $\hat{T}(\underset{\sim}{X}, \underset{\sim}{Y})$ since it is now important to indicate exactly how changes in $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ affect the test.

In Van Eeden [31] the coefficients of the adaptive rank test were monotonized in the construction of the R-estimate. This was done in order to use the results of Hodges and Lehmann [12].

40

The monotonization of the coefficients guarantees that the rank test will be an increasing function of e where e is the displacement of the second sample from the first. This property is used to transfer statements concerning the test statistic to ones involving the location estimate which is helpful in asymptotic investigations.

The coefficients of the adaptive rank test $\hat{T}$ developed here are not monotonized. We will show that for the distributions under which Van Eeden's test is asymptotically fully efficient, the coefficients of the rank test proposed here will be monotonic for all sufficiently large N with probability one.

By not arbitrarily imposing monotonicity, we gain an important advantage with respect to the estimate of Van Eeden in terms of robustness for the following reasons. All long-tailed distributions (e.g., the Cauchy) have redescending score functions, and therefore the LMPRTs for long-tailed distributions have non-monotonic coefficients. Considerable attention is given to long-tailed distributions in robust estimation, and therefore the monotonization of the coefficients should be avoided since it impairs the performance of the estimate for distributions of this type. In particular, imposed monotonization prevents estimates such as Van Eeden's from having asymptotic full efficiency for distributions with redescending score functions. The following lemma allows us to use the results of Hodges and Lehmann for the same distributions for which Van Eeden's estimate is asymptotically fully efficient, while permitting some flexibility for estimation in long-tailed situations.

41

LEMMA 7. Under conditions 1) through 5) of Section 3.1, if $L(\lambda_i)$ is strictly increasing with the increasing λ_i, then $L^*(\lambda_i)$ will also be strictly increasing in i for all sufficiently large N with probability one.

PROOF. From Lemma 1 of Section 2.5 we have that $\hat{L}(u)$ converges to $L(u)$ w.p. 1. By convergence with probability one we have that for all $\epsilon > 0$

$$(4.1.2) \quad \text{Prob}\{|\hat{L}_N(\lambda_j) - L(\lambda_j)| > \epsilon \quad \text{infinitely often} \quad (i.o.)\} = 0 \quad .$$

Let $0 < m = \min\{L(\lambda_i) - L(\lambda_{i-1})\}$, which is greater than zero by assumption. Then

$$\text{Prob}\{\hat{L}_N(\lambda_i) < \hat{L}_N(\lambda_{i-1}) i.o.\}$$

$$= \text{Prob}\{[\hat{L}_N(\lambda_i) - L(\lambda_i)] + [L(\lambda_i) - L(\lambda_{i-1})] + [L(\lambda_{i-1}) - \hat{L}_N(\lambda_{i-1})]$$

$$< 0 \ i.o.\}$$

$$\leq \text{Prob}\{[\hat{L}_N(\lambda_i) - L(\lambda_i)] < -m/2 \ i.o.\}$$

$$+ \text{Prob}\{[\hat{L}_N(\lambda_{i-1}) - L(\lambda_{i-1})] > -m/2 \ i.o.\}$$

$$= 0 \quad . \quad \text{Q.E.D.}$$

We note that the subtraction of the average $\bar{L}_N$ clearly does not affect monotonicity.

We now explicitly define the proposed adaptive estimate of location, or adaptive R-estimate. We have that

42

$$\hat{T}(\underset{\sim}{X},\underset{\sim}{Y}-e) = \sum_{i=1}^{k+1} L^*(\lambda_i) \left\{ \sum_{j=P(i-1)+1}^{P(i)} (\omega_i(-e)) \right\} \quad .$$

Now define

$$e_N^* = \sup\{e : \hat{T}(\underset{\sim}{X},\underset{\sim}{Y}-e) > 0\} \quad ,$$

$$e_N^{**} = \inf\{e : \hat{T}(\underset{\sim}{X},\underset{\sim}{Y}-e) < 0\} \quad ,$$

and finally let

(4.1.3) $$\hat{e}_N = e_N^* + e_N^{**}/2 \quad .$$

We shall only consider the asymptotic theory for the estimator $\hat{e}_N$ in the genuine two-sample case since the theory transfers immediately to the one-sample case.

We now give some results that follow immediately from the work of Hodges and Lehmann, and Van Eeden. The determination of the asymptotic distribution of $\hat{e}_N$ will follow directly from these results.

LEMMA 8. For N sufficiently large, $\hat{T}(\underset{\sim}{X},\underset{\sim}{Y}+e)$ is a non-decreasing function of e for all $\underset{\sim}{X}$ and $\underset{\sim}{Y}$, provided $L(\lambda_i)$ is an increasing function of i.

PROOF. By Lemma 7, with probability one there will be some N_0 such that for all $N > N_0$, $L^*(\lambda_i) < L^*(\lambda_{i+1})$ $i = 0,1,\ldots,k$. Without loss of generality we can assume that $e > 0$. Let $\{\omega_i(e)\}$ denote the w-vector created by adding e to every Y_j. Then $\{\omega_i(e)\}$ will have the 1's in the same or higher positions than the 1's in $\{\omega_i(0)\}$, since the second sample will be shifted in the positive direction. Hence

43

$\hat{T}(\underset{\sim}{X},\underset{\sim}{Y}+e)$ is the sum of the same or larger $L^*(\lambda_i)$'s than those contributing to $\hat{T}(\underset{\sim}{X},\underset{\sim}{Y})$. Therefore $\hat{T}(\underset{\sim}{X},\underset{\sim}{Y}+e)$ is a non-decreasing function of e for all $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ with probability one for all sufficiently large N. Q.E.D.

LEMMA 9. For $\hat{T}$ given by (4.1.1)

$$\hat{T}(\underset{\sim}{X},\underset{\sim}{Y}) + \hat{T}(\underset{\sim}{Y},\underset{\sim}{X}) = 0 \quad .$$

PROOF. By (4.1.1),

$$\hat{T}(\underset{\sim}{X},\underset{\sim}{Y}) + \hat{T}(\underset{\sim}{Y},\underset{\sim}{X}) = \sum_{i=1}^{k+1} L^*(\lambda_i) \left(\sum_{j=P(i-1)+1}^{P(i)} \omega_i \right)$$

$$+ \sum_{i=1}^{k+1} L^*(\lambda_i) \left(\sum_{j=P(i-1)+1}^{P(i)} 1-\omega_i \right) \quad ,$$

$$= \sum_{i=1}^{k+1} L^*(\lambda_i) (P(i) - P(i-1) - 1)$$

$$= 0 \quad \text{by} \quad (3.1.1) \quad ,$$

since $\omega_i(\underset{\sim}{X},\underset{\sim}{Y}) = 1 - \omega_i(\underset{\sim}{Y},\underset{\sim}{X})$, and $L^*(\lambda_i)$ is a symmetric function of the $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ samples. Q.E.D.

LEMMA 10. Under the assumption that $L(\lambda_i)$ is increasing in i, the distribution of $\hat{e}_N$ is (absolutely) continuous provided F is (absolutely) continuous.

PROOF. The result follows directly from Theorem 1 of Hodges and Lehmann [12]. Q.E.D.

LEMMA 11. Under the assumption that $L(\lambda_i)$ is an increasing function of i, the distribution of the estimate $\hat{e}_N$ is symmetric about e, the difference in location between $\underset{\sim}{X}$ and $\underset{\sim}{Y}$.

PROOF. By Lemma 9, $\hat{T}(\underset{\sim}{X},\underset{\sim}{Y}) + \hat{\hat{T}}(\underset{\sim}{X},\underset{\sim}{Y}) = 0$. This is condition (3.3) in Hodges and Lehmann [12] for $\mu = 0$. We showed in Section 2.7 that

$$\hat{T}(\underset{\sim}{X}+e,\ \underset{\sim}{Y}+e) = \hat{T}(\underset{\sim}{X},\underset{\sim}{Y}) \quad .$$

Therefore, we have that $\hat{e}_N$ is symmetric about e by Theorem 2 of Hodges and Lehmann [12].   Q.E.D.

LEMMA 12. Under the assumption that $L(\lambda_i)$ is increasing in i, we have for any real number $e > 0$

$$\text{Prob}\{\hat{T}(\underset{\sim}{X},\underset{\sim}{Y}-e) < 0\} \leq \text{Prob}\{\hat{e}_N < e\} \leq \text{Prob}\{\hat{T}(\underset{\sim}{X},\underset{\sim}{Y}-e) \leq 0\} \quad .$$

PROOF. This result follows directly from Lemma 4 of Hodges and Lehmann [12].   Q.E.D.

We are now ready to give the asymptotic distribution of the proposed estimate.

THEOREM 4. Under conditions 1) through 5) of Section 3.1, assuming the $L(\lambda_i)$ are monotone, we have

$$N^{\frac{1}{2}}(\hat{e}_N - e)$$

converges in distribution to $N(0,V)$ where

$$V = \frac{4 \int_0^1 L(u)^2 \, du}{\{\int_0^1 L(u) J(u) \, du\}^2} \quad ,$$

where $J(u)$ is given by $-f'(F^{-1}(u))/f(F^{-1}(u))$.

PROOF. Because of equivariance, we may let e, the difference in location between the two samples, equal 0 without loss of generality. We have shown in the Corollary to Theorem 2 that for $d_N = 2b/(N \, I(f))^{\frac{1}{2}}$

(4.1.4) $\quad \lim_{N \to \infty} \text{Prob}_{d_N} \{(1/\sigma)[T(X,Y) - \mu] \le u\} = \Phi(u) \quad ,$

where $\text{Prob}_{d_N} \{\ \}$ indicates the probability is evaluated under the condition of a difference of d_N in the two samples, and

$$\mu = \{bN^{\frac{1}{2}}/(2 \, I(f)^{\frac{1}{2}}\} \int_0^1 L(u) J(u) \, du \quad ,$$

$$\sigma^2 = (1/4)N \int_0^1 L(u)^2 \, du \quad .$$

We have

$$P = \lim_{N \to \infty} \text{Prob}_0 \{N^{\frac{1}{2}} e_N \le t\} = \lim_{N \to \infty} \text{Prob}_0 \{\hat{T}(\underset{\sim}{X}, \underset{\sim}{Y} - [t/N^{\frac{1}{2}}]) \le 0\}$$

by Lemmas 10 and 12. Now let

$$b = t[I(f)^{\frac{1}{2}}]/2 \quad .$$

Then

$$P = \lim_{N \to \infty} \text{Prob}_0 \{\hat{T}(\underset{\sim}{X}, \underset{\sim}{Y} - [2b/[N \, I(f)]^{\frac{1}{2}}]) \le 0\}$$

$$= \lim_{N \to \infty} \text{Prob}_{d_N} \{\hat{T}(\underset{\sim}{X}, \underset{\sim}{Y}) \le 0\}$$

46

$$= \lim_{N \to \infty} \text{Prob}_{d_N} \{(1/\sigma)[\hat{T}(\underset{\sim}{X},\underset{\sim}{Y}) - \mu] \leq (1/\sigma)\mu\}$$

$$= \Phi\left[b \frac{\int_0^1 L(u)J(u)du}{[\int_0^1 L(u)^2 du]^{\frac{1}{2}} I(f)^{\frac{1}{2}}}\right]$$

$$= \Phi\left[(t/2) \frac{\int_0^1 L(u)J(u)du}{[\int_0^1 L(u)^2 du]^{\frac{1}{2}}}\right] \quad .$$

This is equivalent to the statement that

$$N^{\frac{1}{2}} \hat{e}_N$$

converges in distribution to $N(0,V)$ where

$$V = \frac{4 \int_0^1 L(u)^2 du}{\{\int_0^1 L(u)J(u)du\}^2} \quad . \quad \text{Q.E.D.}$$

4.2. Description of Monte Carlo Procedures

It is important in simulations to increase the effective sample size of a monte carlo calculation by employing certain variance-reduction techniques. We shall use one of the best known of these techniques, which was prominantly featured in the Princeton Robustness Study, see Andrews, et al. [2]. This method has come to be called the Princeton Swindle. Descriptions of this procedure appear in Andrews et al. [2] and Gross [9], and perhaps the most detailed treatment may be found in Simon [26].

We will now describe the generation of random numbers from the various distributions needed. To implement the variance

reduction procedure we must always represent these observations as normal random variables divided by independent non-negative random variables. The uniform random number generator used was of the usual modulo overflow type. The particular generator chosen produces eight significant digits and proceeds through all 10^8 numbers before repeating. To generate normal random numbers, we generate a bivariate normal pair using polar coordinates and transform to Cartesian coordinates. This is a frequently employed algorithm which conveniently generates two independent univariate normals from two independent uniforms.

To generate random numbers which have a Laplace distribution and which have a representation as a normal divided by an independent random variable, we follow Andrews and Mallows [3]. If $1/(2V^2)$ has an exponential distribution, then a normal random variable divided by V will have a Laplace distribution. Therefore take a uniformly distributed random variable u and let $V = 1/\sqrt{(-2\ln(u))}$.

To generate random variables which have a logistic distribution, we again follow Andrews and Mallows [3]. If K has the distribution of the Kolmogorov distance statistic, then if Z has a standard normal distribution, $Z/(1/2K)$ has a logistic distribution. To generate K we use the fact that $2K^2 = \Sigma_{i=1}^{\infty} W_i/i^2$ where the W_i are independent exponential variables.

The generation of random variables with the normal divided by independent representation is straight-forward for the Cauchy, slash, contaminated slashes, contaminated normals, and wild normals.

See Bell [4], or Andrews, et al. [2] for discussions of the above distributions.

In order to improve the performance of the random number generator we make use of a random scrambler. An array composed of 100 cells is used to store a random number until the cell containing that number is randomly addressed.

In order to estimate the variance of the estimates of variance arrived at in the simulations, summary statistics were printed out every 500 repetitions. The variance of these summary statistics were then used to estimate the variance of the estimates of variance. This is a device commonly used in performing monte-carlo simulations.

4.3. Improving Small Sample Performance

The estimate of location described above was modified slightly in several ways in order to improve small sample performance.

At the suggestion of John Tukey the coefficients of the rank test were smoothed so that the coefficients of consecutive ω_i's were forced to be relatively close in value. It was hoped that this smoothing would decrease the small sample variance of the location estimate. Two smoothing procedures for the estimate of $L(u)$ were investigated. $L^*(u)$ is a step function on $[0,1]$, the steps occurring at $i/N+1$, for $i = 1, \ldots, N$. Define $1_i = (\lambda_{i-1} + \lambda_i)/2$ for $i = 1, \ldots, N$. Thus a smoothed version of $L^*(u)$ for $1_i \leq u < 1_{i+1}$, $0 < i \leq N$, is defined as the line connecting $L^*(1_i)$ and $L^*(1_{i+1})$. The second method was similar to the first except that the smoothing was only

done for $\lambda_1 \leq u \leq \lambda_N$. This second method out-performed the first, and both produced substantial improvement over the non-smoothed version. We note that the smoothed versions do not converge to Gastwirth's asymptotically most powerful grouped rank test.

The necessary symmetry of the coefficients due to the underlying symmetry of the parent distribution may be exploited in the estimation of these coefficients. Thus the estimates of the coefficients of ω_i and ω_{2n+1-i} may be averaged and the average substituted for each of the two estimates. This slightly improves small-sample performance. This procedure may also be regarded as a smoothing device.

Smoothing of the estimate $\hat{G}$, of G was also investigated. The procedure which was finally selected used

$$\hat{G}(u) = (1/2) \, \hat{G}(i/N+1) + (1/2)\hat{G}((i+1)/(N+1)) \quad ,$$

for $i/(N+1) \leq u < (i+1)/(N+1)$.

We also implemented the normal kernel nearest neighbor method to smooth the estimate of f(u). (See Moore and Yackel [21].) This yields

$$\hat{f}_n(u) = (i/n \, R(n)) \sum_{i=1}^{n} K[(u - x_i)/R(n)] \quad ,$$

where K(z) takes the form of a normal kernel.

Note: the ordinary nearest neighbor estimate corresponds to the uniform kernel

$$K(z) = \begin{cases} 1 & \text{if} \quad [z] < 1 \\ 0 & \text{otherwise} \end{cases} .$$

The results of Devroye and Wagner [6] concerning the convergence properties of nearest neighbor density estimates apply to a general class of kernels which includes both the normal and uniform kernels.

It was necessary, finally, to search empirically for combinations of parameter values for the estimator which in some sense maximize its performance on long and short-tailed distributions. The objective was to simultaneously achieve high efficiency for the normal, the 25% contaminated normal (with the slash, denoted .25 1/U), and the Cauchy. This criterion is similar to Tukey's concept of tri-efficiency. The efficiency for the normal was considered to be of greater importance than the efficiencies for the other two distributions. Also, performance relative to the other adaptive estimators in the literature was a consideration.

In our efforts to optimize performance, we needed to determine how many fractiles to use, which fractiles to use, and how many nearest neighbors to use in the density estimate. We soon determined that having more than six fractiles led to poor performance. The optimal number of nearest neighbors for a sample of size 20 was determined to be 11, and for a sample of size 40 the best number was 19. These numbers were relatively insensitive to the choice of the fractiles. The choice of the fractiles was much more crucial. In the investigation of the performance for long-tailed distributions, it was found that by taking the fractiles to be

.30, .50, .70, 1.0, we could nearly equal the performance of the median for those distributions, but that the efficiency for the normal then fell below 80%. It was found that in order to attain 80% efficiency for the normal the fractiles .15, .50, .85, 1.0 sufficed.

One of the most unpleasant features of the computation of R-estimators is the necessity of resorting the combined sample every time a new value is attempted in the minimization procedure. In an attempt to mitigate this difficulty, various one-step procedures using Taylor series were tried, as well as certain extrapolation methods. None of these were found to be satisfactory.

Therefore, the minimization was ultimately performed as follows. The interval from the first to the third quartile was divided into 100 small intervals by 99 equally spaced points. The test statistic (4.1.1) was evaluated at each of the 99 points as well as at the two quartiles. The point at which this test statistic achieved its minimum absolute value was taken to be $\hat{e}_N$. If this value was achieved by more than one point, that point closest to the median was taken as $\hat{e}_N$. This procedure uses the sample quantiles in such a way as to insure location and scale invariance for $\hat{e}_N$.

4.4. Monte Carlo Results for the Adaptive R-Estimate

We present simulation results for the proposed one-sample R-estimate and make comparisons with other adaptive estimators in Table 2 below. The location estimates considered are those proposed by Johns [17], Hodges and Lehmann [12], Stone [28], and Takeuchi [30].

52

TABLE 2

RESULTS OF SIMULATIONS OF R-ESTIMATE

Fractiles = (0, .15, .50, .85, 1.0)

Sample Size = 20

Variance × 20

Standard Deviations in Parentheses

Dist	Reps	$\hat{e}_N$		JOH	H-L	STONE		TAK	"BEST"
Normal	5K/5K	1.200	(.006)	1.137	1.063	1.171	(.005)	1.050	1.000
.25 1/U	5K/5K	1.64	(.008)	1.65	1.70	1.63	(.01)	1.60	1.53
Cauchy	20K/16K	3.2	(.04)	2.8	4.2	2.7	(.02)	3.5	2.3

Sample Size = 40

Variance × 40

Standard Deviations in Parentheses

Dist	Reps	$\hat{e}_N$		JOH	H-L	STONE		TAK	"BEST"
Normal	5K/5K	1.145	(.004)	1.101	1.055	1.094	(.003)	1.036	1.000
.25 1/U	5K/5K	1.59	(.01)	1.60	1.62	1.58	(.008)	1.52	1.46
Cauchy	15K/20K	3.1	(.03)	2.4	3.3	2.5	(.02)	2.6	2.1

53

These are designated JOH, H-L, STONE, and TAK, respectively. The
column labeled "BEST" lists the best performance for each distribu-
tion produced by any of the 65 estimates studied by Andrews, et al.
[2]. The results for the proposed estimate $\hat{e}_N$ and the estimate
proposed by Stone [28] were determined by monte carlo simulation.
The results for the other estimates are taken from Andrews, et al.
[2]. The first column lists the distribution. The second column
lists the number of repetitions for the simulations for the proposed
estimate over the number performed for Stone's estimate. For each
of the estimates considered the estimated variance times the sample
size is given. Standard deviations are given for the estimates for
$\hat{e}_N$ and Stone's estimate.

4.5. Comparison of Results With Other Adaptive Estimates

The study of robust estimation of a location parameter has
concentrated on three classes of estimates; M-, L-, and R-estimates.
R-estimates have been discussed earlier in the present paper. Huber
[14] developed the pure M-estimators which are solutions to equations
of the form

$$\sum_{i=1}^{n} \psi[x_i - M] = 0 \quad ,$$

as well as certain modifications introduced to ensure scale
invariance. L-estimates are linear combinations of order statistics.
It was of interest to compare the estimate proposed here with the
Hodges-Lehmann estimate since the latter is the best known R-estimate

and the only one included in the Princeton Robustness Study (Andrews, et al. [2]). It was also appropriate to compare the proposed estimate with other nearly fully asymptotically efficient adaptive procedures. It should be noted that such procedures do not become leading contenders in the area of robust estimation of a location parameter until sample sizes are greater than 40 (see Hogg [13]).

R-estimates are inherently scale invariant, a feature that is not shared by the most thoroughly studied family of robust estimates, namely, the pure M-estimates. The proper method for obtaining scale invariance is a bothersome question in M-estimation. If R-estimation can be shown to be viable, then this inherent advantage might encourage investigators to return to the study and improvement of R-estimators. On the other hand, if the price paid for this invariance is too high, then the recent lack of interest in this area would appear to be justified.

Since we are to focus our attention on nearly fully asymptotically efficient procedures, we shall henceforth call such procedures ultra-adaptive to distinguish them from estimates which are adaptive for only a specific parameter and make no attempt at full asymptotic efficiency. For instance, M-estimates are all adaptive to some degree as a consequence of the behavior of the method used to obtain scale invariance. In order to determine which estimates to compare with the one proposed here, we need to examine those M- and L-estimators which attempt to gain sufficient information about the distribution to form an asymptotically optimal estimate.

In L-estimation, at least two moderately successful attempts
to accomplish this goal have been made. Johns [17], and Takeuchi
[30] both developed estimators which estimate the coefficients of
an L-estimate by methods which minimize an estimate of the variance.
Stone [28], has proposed an estimator similar to an M-estimator
which involves an estimate of the optimal ψ in the relation

$$\sum_{i=1}^{n} \psi[(x_i - M)/S] = 0 \quad ,$$

where S is a suitably chosen scale-invariant statistic. None of the
other estimators we examined which attempted to achieve at least
nearly full asymptotic efficiency were supported by small sample
results.

Examination of the above simulations shows that the adaptive
R-estimate developed here performs less well than those of Stone,
Takeuchi, and Johns. The proposed estimator compares well with
Takeuchi on 20 data points; and on distributions with longer tails
than the normal, it appears to dominate the Hodges-Lehmann estimator.

It was, however, nearly competitive which is perhaps
surprising considering the general opinion about the sample size
necessary for the successful application of the approach used here
(viz. Huber [15], Hogg [13], and Hajek [10]). The limitation
of the number of fractiles to be less than five did not allow us to
compete very well on both long-tailed distributions and the normal,
because it is difficult to pick the middle fractiles so as to give
good over-all performance. If these are moved from .15 and .85 to

.30 and .70, then the long-tailed distributions are accommodated since $\hat{L}_N(u)$ is then constant over $[0,3)$ and $(.7, 1.0]$ causing $\hat{e}_N$ to behave somewhat like the median. If we move the middle fractiles toward 0 and 1, then the sharp rise in $J(u)$ necessary for good performance on short-tailed distributions may occur, but at the expense of possibly inaccurate density estimation in the tails of long-tailed distributions. This may cause the coefficients of the rank test to be too large for u near zero and one. This discussion may seem pessimistic with regard to the possibility of efficiently using R-estimates in the problem of location estimation; but, in fact, the performance of the proposed estimator is rather better than many investigators would have thought possible. It is not unreasonable to believe that better smoothing techniques or density estimation procedures might further improve the small-sample variances of such R-estimators.

4.6. Summary

In conclusion, we recapitulate the points which have been made concerning adaptive R-estimation. First, an adaptive R-estimate has been constructed which is certainly a reasonably robust estimate of location. The proposed estimate out-performs most previous expectations for an estimator of this type, although there exists estimators already in the literature which are clearly superior to the proposed R-estimator. An explanation of this fact probably lies in the loss of information for small samples that accompanies the use of ranks. Also, estimation in the tails is

crucial; and on samples of size 20 and 40, it seems that it is not a simple matter to estimate the density sufficiently accurately in the tails for small samples.

It may be that the future of R-estimation lies in a different type of adaptation. A possible approach, for example, would be to consider a family of score functions indexed by a parameter p. One could then approximate the "best" p by minimizing an estimate of the asymptotic variance of the R-estimate as a function of p. There is some reason to hope that the performance of such an adaptive estimate could approach that of the leading contenders in the area of robust estimation.

REFERENCES

[1] Albers, W. (1979). Asymptotic deficiencies of one-sample rank
     tests under restricted adaptation. Ann. Statist. 7, 944-
     954.

[2] Andrews, D.F., et al. (1972). Robust Estimates of Location.
     Princeton University Press: Princeton, New Jersey.

[3] Andrews, D.F. and Mallows, C.L. (1974). Scale Mixtures of
     normal distributions. J. Roy. Statist. Soc. B 36, 99-102.

[4] Bell, R. (1980). An Adaptive Choice of the Scale Parameter
     for M-Estimators. Ph.D. Thesis, Stanford University.

[5] Beran, R. (1974). Asymptotically efficient adaptive rank
     estimates in location models. Ann. Statist. 2, 63-74.

[6] Devroye, L.P. and Wagner, T.J. (1980). The strong uniform
     consistency of kernel density estimates. In Multivariate
     Analysis VI, ed., P. R. Krishnaiah.

[7] Fix, E. and Hodges, J.L. (1951). Nonparametric Discrimination:
     Consistency Properties. USAF School of Aviation Medicine,
     Rep. 4, Proj. 21-49-004.

[8] Gastwirth, J.L. (1966). On robust procedures. J. Amer.
     Statist. Assoc. 61, 929-948.

[9] Gross, A.M. (1973). A monte carlo swindle for estimators of
     location. Appl. Statist. 22, 347-353.

[10] Hajek, J. (1962). Asymptotically most powerful rank-order
     tests. Ann. Math. Statist. 33, 1124-1147.

[11] Hajek, J. and Sidak, Z. (1967). Theory of Rank Tests.
     Academic Press: New York.

[12] Hodges, J.L., Jr. and Lehmann, E.L. (1963). Estimates of
     location based on rank tests. Ann. Math. Statist. 34,
     598-611.

[13] Hogg, R.V. (1974). Adaptive robust procedures. J. Amer.
     Statist. Assoc. 69, 909-927.

[14] Huber, P.J. (1964). Robust estimation of a location
     parameter. Ann. Math. Statist. 35, 73-101.

[15]   Huber, P.J. (1972). Robust statistics: A review. <u>Ann. Math.</u>
        <u>Statist.</u> 43, 1041-1067

[16]   Huber, P.J. (1977). <u>Robust Statistical Procedures.</u> Society
        for Industrial and Applied Mathematics. Philadelphia.

[17]   Johns, M.V. (1974). Nonparametric estimation of location.
        <u>J. Amer. Statist. Assoc.</u> 69, 453-460.

[18]   Kiefer, J. (1972). Iterated logarithm analogues for sample
        quantiles when pn ↑ 0. <u>Proc. Sixth Berkeley Symposium on</u>
        <u>Math. Statist. and Prob.</u> 1, 227-244.

[19]   Loftsgaarden, D.O. and Quesenberry, C.P. (1965). A non-
        parametric estimate of a multivariate density function.
        <u>Ann. Math. Statist.</u> 36, 1049-1051.

[20]   Moore, D.S. and Yackel, J.W. (1977). Consistency properties
        of nearest neighbor density function estimators. <u>Ann.</u>
        <u>Statist.</u> 5, 143-154.

[21]   Moore, D.S. and Yackel, J.W. (1977). <u>Statistical Decision</u>
        <u>Theory and Related Topics, II.</u> eds., S. Gupta and D.
        Moore. Academic Press: New York.

[22]   Parzen, E. (1962). On the estimation of a probability density
        function and the mode. <u>Ann. Math. Statist.</u> 33, 1065-1076.

[23]   Parzen, E. (1979). Nonparametric statistical data modeling.
        <u>J. Amer. Statist. Assoc.</u> 74, 105-120.

[24]   Rao, C.R. (1973). <u>Linear Statistical Inference and Its</u>
        <u>Applications.</u> Wiley: New York.

[25]   Rosenblatt, M. (1957). Remarks on some nonparametric esti-
        mates of a density function. <u>Ann. Math. Statist.</u> 27,
        832-837.

[26]   Simon, G. (1976). Computer simulation and swindles, with
        applications to estimates of location and dispersion.
        <u>Appl. Statist.</u> 25, 266-274.

[27]   Stein, C. (1956). Efficient nonparametric testing and estim-
        tion. <u>Proc. Third Berkeley Symposium on Math. Statist.</u>
        <u>and Prob.</u> 1, 187-196.

[28]   Stone, C.J. (1975). Adaptive maximum likelihood estimators of
        a location parameter. <u>Ann. Statist.</u> 3, 267-284.

[29]   Switzer, P. (1971). Efficiency robustness of estimators.
        <u>Proc. Sixth Berkeley Symposium on Math. Statist. and
        Prob.</u> <u>1</u>, 283-291.

[30]   Takeuchi, K. (1971). A uniformly asymptotically efficient
        estimator of a location parameter. <u>J. Amer. Statist.
        Assoc.</u> <u>66</u>, 292-301.

[31]   Van Eeden, C. (1970). Efficiency robust estimation of loca-
        tion. <u>Ann. Math. Statist.</u> <u>41</u>, 172-181.

[32]   Wesley, R.A. (1977). Contributions to the Theory of Robust
        Regression. Ph.D. Thesis, Stanford University.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER No. 5	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* AN ADAPTIVE R-ESTIMATE		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Michael Lee Cohen		8. CONTRACT OR GRANT NUMBER(s) ARO DAAG29-79-C-0166
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, California		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office P.O. Box 12211 Research Triangle Park, N.C. 27709		12. REPORT DATE December 15, 1980
		13. NUMBER OF PAGES 61
14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)*		15. SECURITY CLASS. *(of this report)* Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Robust estimation, robust testing, R-estimator, asymptotic efficiency

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

In 1962 Hajek proposed a test for location which was uniformly asymptotically fully efficient over a large class of distributions. Van Eeden subsequently derived the asymptotic theory for the corresponding R-estimate. Many authors have expressed reservations with regard to the small sample performance to be expected from this approach. In contrast to the methods of Hajek and Van Eeden, the procedure proposed here uses the entire data set to estimate the score function of the locally most powerful rank test. Using nearest neighbor density estimation methods, an estimate of the score function for the asymptotically most

powerful grouped rank test is constructed. This function is itself a step function approximation to the score function for the locally most powerful rank test. Large sample distribution and optimality results are obtained for both the adaptive rank test and the corresponding R-estimate of location. Small sample monte-carlo results are provided.

#5